OE FORM 6000, 2/69

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

ERIC ACC. NO.	P.A.	PUBL. DATE	ISSUE
ED 042 070			
CH ACC. NO.			
AA 000 608		2Nov68	RIEJAN71

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒
ERIC REPRODUCTION RELEASE? YES ☐ NO ☒
LEVEL OF AVAILABILITY I ☒ II ☐ III ☐

AUTHOR
Reid, Ethna R.

TITLE
Evaluation of Teacher Training in a Title III Center.

SOURCE CODE	INSTITUTION (SOURCE)
YEX29710	Granite School District, Salt Lake City, Utah
SP. AG. CODE	SPONSORING AGENCY

EDRS PRICE	CONTRACT NO.	GRANT NO.
0.25:2.20		
REPORT NO.		BUREAU NO.

AVAILABILITY

JOURNAL CITATION

DESCRIPTIVE NOTE
42p.

DESCRIPTORS *Reading Instruction; *Instructional Materials; *Educational Programs; *Program Evaluation; *Teacher Behavior

IDENTIFIERS

ABSTRACT This study is a report on a series of exemplary and instructional reading programs conducted by the Exemplary Center for Reading Instruction and designed to improve reading instruction in kindergarten through grade 12. The following topics are included: (1) evaluation of beginning reading programs, including materials selection, materials analysis, and teaching-behavior analysis in the use of the materials, (2) evaluation of teaching behavior as it relates to classroom management, (3) evaluation of the Reading Center's dissemination services, and (4) basic research as a means of evaluating principles underlying instructional strategy in in-service programs. Numerous tables and three appendices give additional information. (CK)

GPO 870-390

ED 042 070

N-X

Invitational Conference on Testing Problems

November 2, 1968

Evaluation of Teacher Training in a Title III Center

Ethna R. Reid

Exemplary Center for Reading Instruction
Granite School District
Salt Lake City, Utah

AA000 608

EDUCATIONAL TESTING SERVICE
Princeton, New Jersey
Berkeley, California
Evanston, Illinois

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Evaluation of Teacher Training in a Title III Center

ETHNA R. REID
Exemplary Center for Reading Instruction
Granite School District
Salt Lake City, Utah

The Exemplary Center for Reading Instruction (ECRI) is striving to improve reading instruction in grades K through 12 through demonstrations of exemplary diagnostic and instructional reading programs; through in-service teacher training for developmental, remedial, and clinical reading teachers; by collecting, cataloguing, and disseminating information on materials, training methods, and research; and by maintaining liaison with regional and national research and development projects and related institutions to establish cooperative ventures in program development and research.

In this paper it is impossible to describe all of these activities in detail. Because of the emphasis on in-service training and dissemination services at the Reading Center, I will discuss the following: first, evaluation of beginning reading programs, including materials selection, materials analysis, and teaching-behavior analysis in the use of the materials; second, evaluation of teaching behavior as it relates to classroom management; third, evaluation of the Reading Center's dissemination services; and fourth, basic research as a means of evaluating principles underlying instructional strategy in in-service programs.

STRATEGY FOR BEGINNING READING PROGRAMS

An evaluation strategy aimed at beginning reading programs was developed under the direction of Dr. Gabriel Della-Piana, Director of the Bureau of Educational Research at the University of Utah and University Coordinator at ECRI.

Selecting Materials

Large school districts regularly face the practical task of selecting text materials for beginning reading programs, installing and monitoring them, and revising or adapting these materials or their manner of use to get maximum effectiveness and efficiency. Typically, a state committee selects a list of text materials from which local districts may choose one or more or a combination for trial and adoption in their own schools. After materials are selected, however, district personnel must find ways of installing the materials so they are most effectively and efficiently used. The personnel shortage for supervisory and training tasks of this sort makes it mandatory that materials installation be simplified as much as possible.

To provide a data base for selecting materials, we conducted a local comparative study of beginning reading programs, following this three-step procedure:

1. Review current literature to find out which programs are likely to be maximally effective for specified goals within a district population and select one or two of the most likely prospects. Determine also which programs operating within a district are most widely used there.

2. Conduct a comparative evaluation study of these programs so that the program selected for further evaluation and development will be that which yields the greatest achievement gains on the largest number of outcome measures for all ability levels.

3. Select the program that rates highest in a comparative evaluation and modify it to maximize its effectiveness and efficiency.

The materials-selection phase of the evaluation makes use of a treatments (McGraw-Hill *Programmed Reading* plus other experimental programs and controls) by levels (three levels of beginning-of-year reading readiness) analysis of variance design. The results of this analysis tell us which programs are yielding the greatest end-of-year achievement for different beginning-of-year readiness levels. For example, the *Programmed Reading* treatment yielded greater achievement than did the controls for pupils in the initially high and middle levels on the *Murphy-Durrell Reading Readiness Analysis* but not for pupils in the low level. No single reading program was found to be either significantly better than all others on all variables or to be

uniquely effective for students of any given level of preinstructional readiness. Yet McGraw-Hill *Programmed Reading* was favored most frequently, primarily for the pupils in the high and middle reading-readiness levels.

Some tests, such as the individually administered *Gilmore Oral Reading Test*, could not be administered to all pupils in all programs because of efficiency considerations. Instead of a treatments-by-levels design we made a comparison of total *Programmed Reading* versus total controls on regressed gain scores. Here we found that the *Programmed Reading* pupils were superior to the controls on oral reading rate, comprehension, and accuracy measures, but the oral reading accuracy differences were not significant.

The major conclusion derived from this phase of evaluation was that *Programmed Reading* was generally better than the programs used by the controls (eight basal reading programs), the basal reading programs reinforced with the Educational Developmental Laboratory's machines, and the EDL *Listen Look Learn* system groups. Since *Programmed Reading* was generally superior, attention was focused on improving its effectiveness. Also, since it was found weakest for the initially low-readiness pupils on most measures and specifically on oral reading accuracy, future evaluation was focused on these relatively weak spots.

Although final decisions on which treatment is most effective should await longitudinal studies, some decisions must be made on the basis of available data.

Analyzing Materials

We have noted that in spite of its general superiority, *Programmed Reading* was generally not favored on the oral reading accuracy and other measures among low-readiness students. The focus of our materials analysis was influenced by this observation. One basic question guided our materials analysis: Under the current conditions of use of the materials by teachers and low-readiness pupils, what word-recognition errors (oral reading inaccuracies) occurred consistently in one book of the series and did not disappear or diminish in later books?

Answering this question involved selecting a small group of first grade children who were low achievers, testing them on all words

introduced in the Primer and at the end of Book 1, and testing for these same words at the end of Books 2 through 10. For example, words introduced in Book 2 would be tested at the end of Book 2 and also at the end of Books 3 through 10. Table 1 gives a sample of one child's responses to words introduced in the Primer and tested at the end of Books 1 through 5.

Table 1 shows that some words are never missed (a, I, man, pan, yes); some are missed each time they are given from Book 1 through Book 5 (ant); some are missed one or more times and then no more (fat, mat, pin), and others are correct at first and then later missed (no, thin, am). Collecting these data from a group of pupils provides the direction for developing supplementary instructional material and teacher programs. For example, consider the word "fast" introduced in Book 1 and tested at the end of Books 1 through 5. A summary of the errors is presented in Table 2.

The most common errors on words introduced in Book 1 and persisting through Book 5 are: that/did, dig/did, fins/fans, fins/fit, fat/fit, him/ham, hat/hit, ing/in, mitt/mint, Mrs./Miss, pants/pant, pat/pant, pant/pats, pants/pats, sand/sad, sad/sat, sit/sat, sting/sing, sat/sit, thin/this.

The following recommendations are based on the type of data presented in Tables 1 and 2:

1. It is probably inefficient to contemplate any program modification or material supplements for words that are missed frequently in Books 1 and 2 and thereafter never missed.

2. It would be wise to analyze the determinants of word errors which persist over a span of four or five books. For example, the most frequent erroneous readings of "fast" were "fat," "fats," or "fit." An analysis of the test material in Book 1 shows that several items requiring discrimination between "fat" and "fast" are presented (for example, a picture of a thin fish and the statement "this fish is fat/fast"). Since the discrimination problem persists, perhaps some supplementary material should be designed without picture clues for pupils missing this item at the end of Book 1. Alternatively, perhaps the pictures should be modified to prompt the correct response.

The reason *Programmed Reading* was not statistically superior in a significant way to other treatments on oral reading accuracy is prob-

Table 1

Record of one Student's Oral Reading Errors on the PROGRAMMED READING *Word List Across Five Testing Periods*

Student's Name											
Date		2-7	2-21	3-13	3-29	4-24					
Book		1	2	3	4	5	6	7	8	9	10
End-of-Book Test	Word										
1	a										
1	am(2)*		the the			man					
1	an		am	ant		am					
1	ant(2)	at hat	pant hands	hat and	and and	and					
2**	ant			and	hant	pant					
1	I										
4	I										
1	fat	has	fit								
6	fat										
1	man										
2	man										
5	man										
1	mat		hap								
1	no		on	on	on						
3	no										
1	pan (3)										
2	pan										
5	pan										
1	pin		pins								
5	pins										
1	tan					tin					
1	thin		him	the	that	this					
2	thin		D. K	that	that						
4	thin				that						
5	thin										
7	thin										
1	yes										

*Number in parentheses refers to the number of times the word appears in the test.
**"Ant" appears in end-of-book test 1 and again in end-of-book test 2. Similar circumstances hold wherever a word is followed by blackened squares within the table (for example, the word "fat" does not appear as a test word a second time until end-of-book test 6).

Table 2

*Error Record for the Word "Fast," Introduced in Book 1 and Tested for at the End of Books 1 through 5**

Book	1	2	3	4	5
Number of Pupils	7	10	12	11	11
Errors	fat (2) fin it (2) fit dish	sang (2) sat	went fats (2) D.K. past	pig fit	fats fin

*McGraw-Hill PROGRAMMED READING

ably because of the discrimination problems caused by the high dependence of *Programmed Reading* materials on picture cues and grammatical sequence as prompts for filling in blanks. These limitations in materials design can be solved by providing supplementary material or by changing the teacher's use of the regular text materials.

Analysis of Teaching Behavior

At the Exemplary Center for Reading Instruction, we are taking three approaches to the analysis of teaching behavior to reveal deficiencies that may account for the lack of superiority of *Programmed Reading* over controls for the low-readiness pupils. These three approaches are: 1. identifying the most and least effective teachers for the low-ability pupils; 2. observing the behavior of these teachers in using the text materials; and 3. developing observational systems for detecting effective pupil-management techniques.

Identifying differences in teacher effectiveness: By the time we began to identify teacher effectiveness, we discarded the control group be-

cause the *Programmed Reading* classes were not poorer than the controls on any variable. Our objective was to maximize the effectiveness of one of the "best" reading programs in use. So we limited our analysis at this stage to distinguishing between teachers in the *Programmed Reading* group with respect to gains of low-ability pupils. The procedure was as follows:

1. Compute a regression equation for the September (Murphy-Durrell) to May (Gilmore Oral Form B) data for all 89 pupils in the low Murphy-Durrell group in *Programmed Reading*.
2. Determine the residual gain score for each pupil (difference between his predicted May score and his actual May score).
3. For each teacher, tally the number of low Murphy-Durrell pupils who fall above the regression line (perform better than predicted), the number who fall below the line, and the number (if any) who fall on the line.
4. Those teachers who have the greater percentage of their low-ability pupils above the line are the ones who are producing the greatest gains for that ability group.

The data for the 12 *Programmed Reading* teachers in our sample are presented in Table 3. How may these data be used? First of all, for confident decisions about such teacher differences we would want to measure and observe over a period of two or three years. Nevertheless, since we must make some decisions concerning program improvement each year, the data can be put to an immediate use. Teachers 4, 6, and 8 have more students achieving scores below the regression line. The regression line is based on the correlation between beginning-of-year readiness scores and May reading achievement for all 12 classes combined. Teachers 7 and 9 have more students above the regression line. Yet if we look more carefully at other data on these classes, we find that teacher 9 has only 2 low-ability pupils as compared with 4 to 11 for other teachers. The teaching methods used by teachers 7 and 8 could be compared, since they taught equal numbers of low-ability pupils whose socioeconomic characteristics were judged to be identical (barring bias in forming classes) by virtue of their attending the same school. This approach identifies teachers producing the greatest percentage of low-ability pupils who score above the regression line. Once these teachers are identified, they can

be observed to determine which characteristics of their behavior are responsible for the differences in effectiveness.

Comparative data on the instructional methods used by teachers 7 and 8 would be useful to guide program improvement. However, it would be more desirable to have data on a larger number of teachers. The ideal procedure would be to begin with a larger number of classes in schools each having three or four first grade classes and where there was random assignment of children to classes.

Table 3

Number of Low Reading Readiness Students Falling Above, On, or Below the Regression Line for Predicting May Vocabulary and Comprehension Scores on the* GATES-MACGINITIE READING TEST

Teacher	School	Group Size	*Above*		*On*		*Below*	
			Vocabulary	Comprehension	Vocabulary	Comprehension	Vocabulary	Comprehension
1	A	10	4	5	—	—	6	5
2	A	9	4	5	—	—	5	4
3	B	7	4	4	—	—	3	3
4	B	11	3	4	1	0	7	7
5	B	7	4	2	—	—	3	5
6	B	8	2	2	—	—	6	6
7	C	6	6	5	—	—	0	1
8	C	6	2	1	—	—	4	5
9	D	2	2	2	—	—	0	0
10	E	9	3	4	—	—	6	5
11	E	4	2	1	—	—	2	3
12	E	10	4	4	—	—	6	6

*Scores ranging from 16-73 on the MURPHY-DURRELL READING READINESS ANALYSIS, September 1968

Observing teaching behavior: The degree to which all teachers used the teacher's guide was observed during the materials-selection phase of evaluation. Teacher 7 (who obtained better-than-average gains for low-ability pupils) deviated more from the teacher's guide and spent more time with low-ability pupils than did teacher 8. Teacher 8 was rated "1" (high fidelity to the guide) and teacher 7 was rated "4" (low fidelity to the guide) on a five-point scale. Both of these teachers moved at the end of the first year's study, so we were not able to observe their behavior on other dimensions. Table 4 lists some of the dimensions of teaching behavior which we see as relevant.

Note that we have included a category for observation of the pupils' behavior, since an adequate description of how a teacher manages the class must include some of these data. More detailed behavioral data on pupils can be obtained by procedures which will be discussed later.

Informal observations made while using the teaching-behavior checklist have identified some variability in teaching behavior including a rather common lack of detailed diagnostic and prescriptive teaching, a defect which should significantly affect the performance of slower pupils. At any rate, behavioral data can be applied toward the goal of maximizing a reading program's effectiveness by identifying behavior characteristics which discriminate between high- and low-gain teachers of low-ability pupils, by obtaining base rates on significant teacher and pupil behavior, and by providing direction for in-service training programs. The base rates, of course, may be used as a baseline against which teaching behavior observed during and following teacher training can be compared.

Teaching-observation systems currently being developed: An example of an observation system for one aspect of teaching behavior listed in Table 4 (*reinforcement contingent on performance*) is presented in appendixes A and B. The system is designed for observing and recording the extent to which teachers establish (promise) contingent stimuli ("If all of you finish your work before 10:30 we will play our spelling game.") and apply them (actually carry out the contingencies promised). The system is broad enough to deal with positive reinforcers, punishment and escape contingencies, and application of contingencies not previously promised.

The system's sensitivity to teaching differences was demonstrated in a pilot study. Five teachers trained in contingency management were compared with 14 untrained teachers observed in a previous study. The five teachers were observed in four separate half-hour

Table 4

Categories for Observation Schedule for Teacher and Pupil Behavior (Low-ability Pupils) in Teaching with McGraw-Hill* PROGRAMMED READING *Materials*

Teacher Behavior

—end-of-book test
  administers
  listens to oral reading
  records errors
    accuracy of recording
  records causes of errors
    accuracy of causes

—teacher time on task
  percent of reading time teacher not diagnosing, prescribing, or teaching

—making prescriptions of objectives (written or mental)
  related to diagnostic data
  specific response described
  situation in which response is specified
  criteria for an acceptable response specified

—prescriptive teaching (described/conducted)
  achievable bits
  prompts for evocation of response
  feedback
  fade prompts
  overlearning
  varied context practice
  related to diagnosis
  reinforcement contingent on performance

—fidelity to teacher's guide (to be listed in detail)

Pupil Behavior

—number of pupils in class/number with *Murphy-Durrell* score 73 or below

—each child: workbook number/date

—each child: total time allocated to reading period/percent time at task

—each child: time allocated to independent reading/percent time at task

---

*Instructions for time sampling of behavior not included

sessions at the beginning of the reading period. The results indicated that:

1. Verbal positive contingent reinforcement was more frequent for trained teachers (28 per half-hour) than for untrained teachers (11 per half-hour).
2. Verbal negative contingent reinforcement was more frequent for trained teachers (12 per half-hour) than for untrained teachers (3 per half-hour).

Thus, teachers may be observed for the categories indicated, differences between high and low gain-producing teachers may be observed, and training programs may be initiated to produce desired changes in teaching behavior.

EVALUATION STRATEGY FOR CLASSROOM MANAGEMENT BEHAVIOR

The current emphasis within the realm of individualized instruction is upon the relationship between student work skills and teaching behavior. Individualized instruction requires that the teacher spend time with each child on his program or his products. The teacher's role becomes that of expert in diagnosis, prescription preparation, and general trouble shooting. Ideally, with no pragmatic demands, this might be met by a one-to-one tutorial setting, at least for much, if not all, academic instruction. Realistically, it must be met by the teacher moving from child to child as needed, yet maintaining a productive classroom. That is, children should be able to work individually, yet get help when their progress is thwarted.

In many classrooms the teacher decides when his pupils should work; he continually prods them to attend, to continue working, and to complete what they are doing. He keeps them moving from subject to subject and from unit to unit. Unfortunately, these procedures are not always effective in maintaining a productive classroom.

The reasons for inadequate development of self-controlled work skills are probably numerous. Ideally, we would like children to continue working without constant teacher intervention. Yet many

teachers quickly interrupt when a student stops working, and ignore him when he is working satisfactorily. These teachers act as if their task is to detect and correct "misbehavior" only. Such differential attention might inadvertently reinforce idleness or misbehavior. For example, Wesley C. Becker, University of Illinois (unpublished), experimentally demonstrated in a classroom that the more the teacher told children who were out of their seats to sit down, the more they left their seats. Teacher reinforcement of other poor work skills is probably also common.

Another problem concerns the kind of reinforcement contingencies teachers use to maintain work skills. In some classes the essential "control" procedure is an "escape" contingency. That is, rather than positively reinforcing appropriate behavior, the teacher bombards his students with repeated instructions, threats, and criticisms when they are not working. The children go to work or do whatever is necessary to terminate, and thus escape, their teacher's unpleasantness. The verbal barrage ceases when the children return to work. Even though it gets students to work, such a procedure is likely to fail in the long run because it develops no motivation to work other than the motivation to escape the noise. When the threats and directions stop, the work may also stop. The teacher is then trapped in a predicament in which he must continually repeat the instructions, threats, and demands if he is to maintain student behavior.

One phase of ECRI's in-service teacher-training programs includes the development of innovative practices in strategically located "Skill and Product Development Classrooms" which serve as exemplary supply depots from which area training and demonstration teachers set up similar programs in other schools.

Practical, easily taught techniques for teachers in establishing and maintaining students' independent work skills constitute the primary objective of one of the SPD classrooms. The observation system for evaluating teaching behavior in the classroom has been developed and is currently being refined under the direction of Dr. Howard N. Sloane, Jr., University of Utah.

Program Preparation and Reinforcement System

Two classes of independent work-skills behavior have been identified. The first concerns paying attention to instructions and not engaging in

competing behaviors while the teacher is giving instructions. The second relates to pupil maintenance of independent work. Specific programs to develop each of these classes of behavior are being used.

A reinforcement system requiring very little teacher effort has been designed. Students can earn points for the class as a whole during classroom activities in which the teacher must interact with several students simultaneously. The teacher tallies these points on an inexpensive, electronic counter which students can see and hear from anywhere in the classroom. Students can earn individual points for correctly completing an academic unit, and they may periodically trade their points for "back-up reinforcers" such as classroom privileges and activities.

Evaluation Design

The major evaluation instrument for student and teacher behavior has been developed, and its reliability is now being checked. Appendix C includes procedures for coding major areas of student and teacher behavior and a summary copy of the classroom behavior scale.

Through the use of this instrument by trained observers, teacher and student behavior changes can be evaluated by observation and rating before and after training sessions. Initial reliability data on the classroom behavior scale are included in Table 5.

Table 5 shows the percent of agreement on the classroom behavior scale among three observers who had been trained over a three-month period. The observers rated five randomly selected second grade students over nine thirty-second observation periods as described in the rating procedures.

Percent agreement was calculated as:

$$\text{Percent agreement} = \frac{\text{number of agreements}}{\text{number of agreements} + \text{number of disagreements}}$$

In assessing agreement and disagreement, different rating codes (0's and I's for example) applied to the same behavior by different raters within a single rating interval constituted disagreement.

Raters indicated that they recognized a behavior's absence by drawing a diagonal line across all rating intervals. The use of a diagonal line left no question as to whether raters were sure the behavior

Table 5

Estimate of Interobserver Agreement on the Classroom Behavior Scale

Behavioral Subclasses	*Percent of Agreement*
AREA 1 SB	
U/D	87
AREA 2 NV-T	
p	100
c	100
h	99
a	100
+	99
—	99
AREA 3 V-T	
+	100
—	99
I	99
O	100
AREA 4 V-O	
+	91
—	91
I	69
O	88
AREA 5 TI	
i	75
g	94
c	66
◯	63
'	89
—	77

in question did or did not occur. Therefore, agreement in using the diagonal line constituted agreement in using the rating codes which it replaced on the scoring records.

Additional evaluation data will consist of records of assignment completion and correctness, measures of student attending behavior, and some general measures of academic achievement. The demonstration class and experimental (or field training) classes will be compared with themselves and with other control classes.

DISSEMINATION EVALUATION

The Exemplary Center for Reading Instruction maintains four distinct dissemination avenues, each requiring independent evaluation. The four include: (a) out-of-Center services such as those rendered by the area training teachers; (b) library loans; (c) demonstrative and exemplary functions within the Center, including visits and tours; and (d) dissemination via the mail service. Evaluation programs for (c) and (d) are in progress under the direction of Dr. Jon E. Atzet, Reading Center psychologist and co-editor of the ECRI *Newsletter*.

Dissemination via the Area Training Teacher

This dissemination medium involves the Skill and Product Development Classrooms described earlier and ECRI's area demonstration training teachers. The area training teachers instruct classroom teachers in individual prescription techniques, in establishing effective independent work skills among their students, and in teaching their students an elemental approach to critical reading. The teachers are the targets; it is their teaching techniques and behavior management that are to be modified and honed to raise classroom efficiency and productivity.

Dissemination via Library Loan

The library faces a unique problem in evaluating its program. Though

its holdings make up several distinct categories or subject areas, its patrons are individuals who frequent the library to fill their particular needs. Several patrons might use and assess the worth of a certain item, but the premium each places upon it will differ from person to person and will be determined by the unique way in which each person uses it. Library holdings, therefore, have no indigenous function and cannot be compared against a success criterion. Service, on the other hand, can be.

Circulation records reflect demand, and demand reflects worth or value. Since functionally similar holdings are categorized together in the library's organization, each category's intrinsic value can be determined by statistical comparisons of circulation records among categories. Further, circulation records for all categories can be combined to establish periodic service records for the library as a whole.

Librarians do not normally tally the number of inquiries they receive about materials their facility does not have available. Yet such inquiries demonstrate an interest in certain materials and can, therefore, be used as evaluative data. At ECRI such inquiries are recorded, categorized, and tallied so that they reflect interest in materials not available through the library, and materials which are frequently requested are later installed in the library.

The library evaluation program has also been directed toward estimating the demand for its services. Requests reach the Center by mail, by telephone, and in person. All request letters are filed; telephone calls come through the front desk and are rerouted to their ultimate destination where their messages are recorded and filed until they are counted and categorized. Requests filled in person are estimated from depleted materials stores and from library circulation records.

During the first quarter (January-March 1968), readers requested 107 copies of articles reviewed in the ECRI *Newsletter*, over 300 *Library Resources* books were sold, visitors took away tens of thousands of the many teacher aids and in-service training pamphlets and bulletins, and library circulation reached 33,713.

Dissemination via Visits to ECRI

Reading Center visitors participate in demonstrations, workshops, and lectures; use the library; consult with teachers and other per-

sonnel; tour the Center; and participate in ECRI-sponsored functions held outside the Center.

Data from the questionnaire that we circulate among visitors reveal that during the first quarter (January-March 1968), approximately 59 percent of the Reading Center visitors were teachers, 15 percent were educational administrators, and 26 percent were from other occupations ranging from university students to commercial welders. Seventy-five percent of the visitors came from within the Rocky Mountain region and 25 percent came from outlying states.

The reasons which Reading Center visitors listed for their visits indicate that 6.5 percent came because their children were being instructed in the Reading Clinic, 42.5 percent came to participate in either demonstrations or workshops, 5.9 percent came to use or learn about the Reading Center library, 48.7 percent came to tour the Reading Center or for a general introduction to its facilities and functions, and 2.5 percent came for unstated reasons. Several people came for a variety of reasons and were therefore included in more than one tally.

ECRI's influence extends far beyond the Rocky Mountain region. It has served all of the continental United States, Alaska, Hawaii, and parts of Canada. Workshop participants, consultants, and visitors to ECRI have represented 29 states including Hawaii and Alaska.

Dissemination via the Mail Service

Through the mail service, ECRI dispenses such items as newsletters, bulletins, professional reports, and announcements. Evaluating this medium can be difficult because there is no personal contact between the disseminating and the target agencies. If readers are to evaluate the mailings they have received, a follow-up effort must be made to reach them.

A follow-up evaluation program is expensive. Besides the cost of two-way postage, a self-explanatory, mail-sized evaluation form would have to be designed, mailed, returned, and sorted, and its contents tallied, all of which would consume many costly man-hours.

A follow-up evaluation program makes additional demands upon the evaluator. He is asked to evaluate materials he read some time ago. Provided the reading material has not been misplaced or discarded, the evaluator must refresh his memory on pertinent points by

rereading and pondering the information in the light of the questions on the evaluation form. Then he must complete the evaluation form and return it.

Many potential evaluators habitually shun evaluation programs because of earlier experiences with demanding questionnaires. Others faithfully comply by filling out an evaluation form but neglect the more important task of preparing themselves to do so. Thus they sabotage evaluation accuracy and utility.

Follow-up evaluation programs are often weak because of insufficient compliance. In instances of indirect confrontation such as the follow-up evaluation, compliance is generally inversely proportional to the effort demanded by the evaluation questionnaire unless it is controlled by an attractive form of reward. But in this case the reward carries an intangible, and too often valueless, "do-it-for-science" flavor.

Evaluation relevancy often suffers because potential evaluators are not adequately qualified. Comparison is a fundamental part of evaluation. Its application is exemplified in the before-after and the experimental-control techniques used in scientific investigations. The need for comparison in evaluation imposes the qualification that an evaluator must be well acquainted with the material—and similar materials—he is to appraise. The more extensive his familiarity with related materials, the better equipped he is as an evaluator.

If comparison in evaluation is infeasible, as might be the case with certain innovations which neither replace nor resemble other methods or materials, more stringent qualifications are demanded of the evaluator. He must be willing to take the time to survey the material carefully, looking for inherent merit, potential alternatives, and potential pitfalls. The evaluation, in this case, must reflect exclusively upon the materials being evaluated.

The ill-equipped evaluator tends to shower his subject with praise, virtue, flattery, and so on, which translate into positive or favorable evaluative data. This "halo effect" is tremendously effective in boosting self-estimations, but such evaluation returns do not reflect upon the actual quality of the program they were intended to assess.

Follow-up evaluation program: An evaluation program should be extensive enough to disclose a program's inadequacies, shortcomings, minor faults, and, of course, its strong points. A program design, free from internal difficulties, will provide more latitude for approaching the evaluation program's purpose.

Most problems encountered in a follow-up evaluation program can be controlled through questionnaire design. First, an evaluation questionnaire should be small enough to be sent in a regular mailing. Mail pieces and evaluation questionnaires of equal size can be sent as a unit, saving half of the postage which must be spent if questionnaire size demands that it be sent separately. Incorporating a questionnaire into a standard mailing minimizes the time lag between reading and evaluating. It alerts the reader-evaluator to read with the evaluation objective in mind. The questionnaire guides his reading and prevents him from having to review in order to appraise materials that he once read. Eliminating the time lag between reading and evaluating increases evaluation validity by reducing memory loss and thus the "halo effect" that is most prevalent under conditions of ignorance and/or failing recall. The greatest benefit, however, is that a reduction in effort can increase compliance.

When a questionnaire accompanies any disseminated material, it becomes feasible to gather evaluative data randomly on each mailing rather than from selected readers at selected times. The selection factor alone can reduce evaluation validity because selection is directly contrary to scientific sampling methods.

Second, the questionnaire should carry only those questions which are most relevant to evaluation. The questions should be concise, terse; none of them should be open-ended. Well-structured questions shorten a questionnaire's length and complexity, thereby expediting both the response to it and the tallying of data from it once it is returned. Structured questions provide the evaluator with an evaluation guide; they prevent him from having to conjure up his own evaluation categories. Structured questions provide the evaluator with a type of reading guide enabling him to read for evaluation as well as for his own purposes.

Questions can be leveled at the evaluator's qualification level or designed to control depth of thought. Where evaluation validity is a major concern, question and questionnaire structure can counterbalance lack of qualification.

Third, an evaluation questionnaire should carry a brief description of its purpose stated in such a way as to emphasize the importance of evaluation and the contribution made by each evaluator. The emphasis should be directed at increasing compliance and conscientiousness of effort.

Evaluating the ECRI *Newsletter:* Of the countless objectives that

could be tied to an educational publication, only those that reflect the publishing organization's expressed purpose and its readers' needs should be considered in establishing the criteria against which the publication is to be evaluated. Even though the readers' needs appear to be the primary concern, they are not so important that they should be allowed to alter the publishing organization's purpose. For example, readers cannot legitimately demand information that lies outside the publisher's domain and complain if they do not get it. From the outset, then, the publishing organization is obliged to state its objectives clearly and publicly, and to serve its readers within the limits established by these objectives. Evaluation is the process by which readers assess, primarily, the degree to which a publication's contents actually reflect its objectives, and, secondarily, the degree to which the publication fills their own needs.

The *Newsletter's* objectives, as formulated by the ECRI staff, were: 1. to disseminate information on the Reading Center's functions; 2. to disseminate reading-research findings derived from studies sponsored by the Reading Center; 3. to provide teachers with effective exemplary practices and classroom aids; 4. to provide educators with a medium for publicly commenting on current practices and innovations in teaching reading; and 5. to help educators keep abreast of changes in teaching techniques, materials, and educational philosophies.

In pursuit of its objectives, each issue of the ECRI *Newsletter* features a progress report on ongoing reading-related research sponsored by the Reading Center (shown in Table 6 as Section A) and carries a synopsis about the author (Section B). A third part of the *Newsletter* (Section C) provides a detailed description of an exemplary teaching practice. Section D is reserved for readers who wish to comment on the *Newsletter*'s content and related issues. Section E reports on ECRI-sponsored projects other than concurrent research. Section F presents a review of pertinent, recent research in reading from throughout the world and offers these reports in their entirety through ECRI upon request. Section G provides short, concise suggestions for increasing motivation to read. Sections H and I refer respectively to the cartoons and photographs which supplement the text.

Designing an evaluation form: An evaluation study was undertaken to determine how effectively the *Newsletter*'s contents are fulfilling its objectives and satisfying its readers. An extensive evaluation questionnaire was developed and repeatedly condensed until it fit on one side of an 8½-inch x 11-inch sheet. The other side of the sheet was divided

horizontally in two. The reasons for the evaluation and the general directions were printed on one half; a self-addressed, postage paid, return cover filled the other. Appendix D contains a sample copy of the evaluation questionnaire. The evaluation form was mailed with an evaluation issue of the ECRI *Newsletter*.

Item I on the questionnaire allowed those on the mailing list to either continue or discontinue their subscriptions by checking the appropriate box and returning the questionnaire. ECRI originally adopted the policy of mailing the *Newsletter* regularly to everyone on its mailing list. This policy was to guarantee all potential readers an opportunity to experience the *Newsletter*'s impact, to develop a personal interest in it, and perhaps to pass a copy on to others who might share their interest and submit requests for the publication. The policy has worked well. More than 350 subscriptions were received from March to June 1968 from readers who were introduced to the *Newsletter* by friends.

Because those who are not educators as well as educational administrators who are far removed from the classroom were also represented in the swelling 7,000-entry mailing list, circulation probably exceeded readership. Their actual interest in the *Newsletter* was probably very low, and because of the disinterest, it was expected that some of them would withdraw their subscriptions.

Item I was included in the evaluation questionnaire also to separate *Newsletter* readers from nonreaders to preclude using nonreaders' data in the evaluation.

Item II was incorporated into the evaluation questionnaire: 1. to measure the relative extent to which each section of the newsletter was read; 2. to determine which sections readers thought should be given more space or emphasized for their benefit; and 3. to isolate those sections which were of little or no value to readers. Such information was to guide the editors in redesigning the *Newsletter*'s format to satisfy its readers' needs more effectively.

Item III represented an effort to generate ordinal data which could be applied toward a minute and exhaustive evaluation of individual *Newsletter* sections. Readers were asked to rate (on a 1-5 scale representing excellent through poor) sections A through I on their relative clarity, informativeness, interest value, importance, utility, applicability, practicality, originality, and influence—a variety of attributes that could be applied indirectly to an assessment of the objectives outlined for the publication as a whole.

The purpose of Item IV was to identify and categorize all occupations represented in the readership. Grouping by readership was to provide the third dimension for the data analysis.

Item V provided evaluators with space to comment freely on the *Newsletter*. Open-ended responses could supply relevant, qualifying information not allowed for elsewhere in the evaluation questionnaire.

Results: The response to Item IV demonstrated that the ECRI mailing list contains a diverse sample of the educational populace, a full cross section of people representing education in one way or another. Several distinct groups emerged from the response sample: elementary school teachers (shown in Tables 9-11 as ET), elementary school administrators (EA), secondary school teachers and administrators (STA), college and university teachers and administrators (CTA), and a fifth group comprised of other administrators and educational specialists (O).

Inasmuch as each of the above groups was thought to have a somewhat different professional mission, it was hypothesized that each would assess the various sections or the entire *Newsletter* differently. Table 6, in presenting the analysis of the response to Item III, shows this hypothesis to be a misconception; GROUPS did not differ significantly among themselves ($p > .10$). Neither were SECTIONS by GROUPS nor RATINGS by GROUPS interactions significant ($p > .25$). These results mean that readership groups did not differ among themselves in the way each of them rated the *Newsletter* sections and used the rating categories.

Collectively, however, groups rated each of the *Newsletter* sections and applied each of the rating categories differently (Table 6; SECTIONS, RATINGS, and SECTIONS by RATINGS; $p < .01$). In Table 7, the *Newsletter* sections are ranked according to the magnitude of the overall rating score received by each. Exemplary Teaching Practices (C) ranked highest; then came Reading Research Review (F), Feature Article (A), Reading Keys (G), ECRI-sponsored Projects (E), Letters to the Editor (D), Photographs (I), Cartoons (H), and About the Author (B). All ordered pairs of rankings, except F and C, E and G, I and D, H and D, and H and I, were significantly different from one another (Table 7).

In Table 8, the rating categories are ranked according to their individual total scores—the sum of all numerical ratings for each rating category contributed by all readership groups across all *News-*

letter sections. The *Newsletter* was rated highest on clarity (12); then came informativeness (11), interest value (10), importance (9), utility potential (8), originality (5), influence (4), practicality (6), and applicability (7), respectively. All ordered pairs of ranking, except 8 and 9, 5 and 9, 5 and 8, 4 and 8, 4 and 5, 6 and 8, 6 and 5, and 6 and 4, were significantly different from one another.

Newsletter sections were further ranked from the data received in Item II according to sections read most regularly (readership strength). The Feature Article was the most heavily read (Table 9). Exemplary Teaching Practices, Reading Research Review, ECRI-sponsored Projects, About the Author, Reading Keys, Letters to the Editor, Photo-

Table 6

Comparison[a] of Preferences for and Ratings of Nine NEWSLETTER *Content Sections by Five Readership Groups*

Variance Source	*df*	*MS*	*F*	*p*
GROUPS	4	35.58	2.25	NS*
ERROR	20	15.78		
SECTIONS	8	35.11	12.36	<.01**
SECTIONS × GROUPS	32	2.64	.92	NS
ERROR	160	2.84		
RATINGS	8	2.62	4.12	<.01**
RATINGS × GROUPS	32	.504	.79	NS
ERROR	160	.636		
SECTIONS × RATINGS	64	.561	2.47	<.01
SECTIONS × RATINGS × GROUPS	256	.224	.98	NS
ERROR	1,280	.227		

[a]The data were analyzed via a 5 × 9 × 9 analysis of variance having repeated measures over the second two dimensions.

*NS = not significant

**Partitions of the variance are presented in Tables 7 and 8.

graphs, and Cartoons followed in that order. Readers agreed that Exemplary Teaching Practices were needed more than any other section (Table 10). Reading Research Review, ECRI-sponsored Projects, Reading Keys, Feature Article, Photographs, Letters to the Editor, Cartoons, and About the Author followed in that order. Readers agreed that the section having least value to them was Letters to the Editor (Table 11). Photographs, Cartoons, Reading Keys,

Table 7

Inter-section Comparison[a] of all Ordered[b] Pairs of NEWSLETTER *Sections*

(Difference Matrix[c])

	C	*F*	*A*	*G*	*E*	*D*	*I*	*H*	*B*
C		3	20*	65**	71**	181**	189**	193**	221**
F			17*	62**	68**	178**	186**	190**	208**
A				45**	51**	161**	169**	173**	191**
G					6	116**	124**	128**	146**
E						110**	118**	122**	140**
D							8	12	30**
I								4	22**
H									18**
B									

[a]The variance was partitioned via the Newman-Keuls Sequential Range Statistic.

[b]NEWSLETTER sections are ranked on both the abscissa and the ordinate in the ascending order of total rating scores.

[c]Any score in the matrix is the absolute difference between the total rating scores assigned to the sections directly opposite it on both the abscissa and the ordinate. The lowest rating indicates highest desirability.

*Denotes significance beyond the .05 level.

**Denotes significance beyond the .01 level.

About the Author, ECRI-sponsored Projects, Reading Research Review and Exemplary Teaching Practices (tied ranks), and the Feature Article followed in order of ascending value.

One of the most significant findings of this evaluation emerged from complementary results produced by two independent analyses: (a) the degree to which the *Newsletter*'s contents reflected its objectives (an analysis of Item III data), and (b) the degree to which the *News-*

Table 8

Inter-category Comparison[a] of all Ordered[b] Pairs of Rating Categories

(Difference Matrix[c])

	12	*11*	*10*	*9*	*8*	*5*	*4*	*6*	*7*
12		36**	49**	59**	64**	64**	68**	70**	82**
11			13**	23**	28**	28**	32**	34**	46**
10				10**	15**	15**	19**	21**	33**
9					5	5	9*	11*	23**
8						0	4	6	18**
5							4	6	18**
4								2	14**
6									12**
7									

[a]The variance was partitioned via the Newman-Keuls Sequential Range Statistic.

[b]Rating categories are ranked on both the abscissa and the ordinate in the ascending order of total rating scores.

[c]Any score in the matrix is the absolute difference between the total rating scores assigned to the rating categories directly opposite it on both the abscissa and the ordinate. The lowest rating indicates highest desirability.

*Denotes significance beyond the .05 level.

**Denotes significance beyond the .01 level.

letter's contents fulfilled its readers' needs (an analysis of Item II data).

Item II data show that readers read most often, requested most often, and valued highest all *Newsletter* sections relating to exemplary classroom practices (Sections A, C, F, and G; Tables 9, 10, 11). Item III data show that readers consistently rated these same sections higher than the rest (Table 7). The fact that, among the rating categories, the *Newsletter* was rated lowest on practicality and applicability (Table 8) can be explained by the significant nonadditive variance (Table 6; SECTIONS by RATINGS) that remained beyond significant SECTIONS and RATINGS main effects. A further analysis of these data shows that Exemplary Teaching Practices, as well as the other sections carrying exemplary teaching practices, were exempted from the low practicality and applicability ratings. Partitioning the nonadditive variance placed Reading Research Review, Exemplary Teaching Practices, Reading Keys, and the Feature Article in a group statistically above the remaining sections insofar as practicality and applicability were concerned.

Table 9

Ranking of* NEWSLETTER *Sections According to Readership Strength*

Readership Groups	NEWSLETTER *Sections* *A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
ET	1	6	2	8	5	3	4	7	9
EA	1	6	2	7	4	3	5	8.5	8.5
STA	1.5	5	3	6.5	4	1.5	8.5	8.5	6.5
CTA	1	4	2.5	7	5.5	2.5	8.5	8.5	5.5
O	1.5	5	1.5	7	4	3	6	8	9
TOTAL*	6.0	26.0	11.0	35.5	22.5	13.0	32.0	40.5	38.5

*Ranks were analyzed via the Friedman two-way analysis of variance, where:
$\chi_r^2 = 33.94$ and: $p < .001$

*The lowest total score identifies the most heavily read section. A χ_r^2 of 33.94 indicates significant differences among all pairs of rankings, i.e., section A is read significantly more than its closest competitor, section C, and so on.

Table 10

Ranking[a] of NEWSLETTER *Sections According to Readership Need*

Readership Groups	NEWSLETTER *Sections* *A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
ET	5.5	9	1	5.5	4	2	3	8	7
EA	5	8	1	6	3.5	2	3.5	7	9
STA	5	5	1	9	3	2	7.5	7.5	5
CTA	3	8	2	8	4	1	5	8	6
O	5	9	1	8	3	2	4	6.5	6.5
TOTAL*	23.5	39.0	6.0	36.5	17.5	9.0	23.0	37.0	33.5

[a]Ranks were analyzed via the Friedman two-way analysis of variance, where:
$\chi_r^2 = 33.94$ and: $p < .001$

*The lowest total score identifies the most needed section. A χ_r^2 of 28.07 indicates significant differences among all pairs of rankings, i.e., section C is significantly more valuable than its closest competitor, section F, and so on.

BASIC RESEARCH

While the Exemplary Reading Center's mission is primarily that of program development, training, and dissemination, it has some commitment to basic research, which is part of the total evaluation program. In-service programs involve instructional strategy. Basic research has been supported because of its focus on basic psychological principles underlying some of the instructional methods used in our in-service programs. Typically, basic research is sponsored by the Reading Center in cooperation with other agencies such as the University of Utah. Two doctoral dissertations* were carried out

*Alter, Madge. Identification of high probability responses and their use as reinforcers. Doctoral dissertation, University of Utah, 1968.
Chan, Adrian. An analysis of Premack's rate differential response theory. Doctoral dissertation, University of Utah, 1968.

under the direction of Drs. Della-Piana and Sloane. Both studies dealt with Premack's theory that of two events (A and B) the one occurring more frequently (A) will reinforce the lower-frequency event (B).

Dr. Alter was concerned with developing procedures for identifying high-probability responses and determining their stability and their utility as reinforcers for low-frequency responses. Premack hypothesizes that if a response occurring at a higher rate is made contingent upon a lower-rate response, the high-rate response can be used to reinforce (increase the frequency of) the low-rate response. If this principle is to find practical application in the classroom, a procedure must be devised whereby teachers can chart response frequencies for commonly occurring classroom activities.

An initial study was designed to do just that. A method was developed for identifying high- and low-frequency activities for individual pupils. Commonly occurring classroom activities were

Table 11

Ranking[a] of NEWSLETTER *Sections According to Perceived Value*

Readership Groups	NEWSLETTER *Sections*								
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
ET	9	4	7.5	2	5.5	7.5	5.5	3	1
EA	7.5	4.5	7.5	2	7.5	7.5	4.5	3	1
STA	7	3	7	3	7	7	7	1	3
CTA	8.5	7	5	1.5	5	8.5	1.5	3	5
O	8	5	6	2.5	8	8	4	2.5	1
TOTAL*	40.0	23.5	33.0	11.0	33.0	38.5	22.5	12.5	11.0

[a]Ranks were analyzed via the Friedman two-way analysis of variance, where:
$\chi_r^2 = 24.64$ and: $p < .01$

*The lowest total score identifies the least valuable section. A χ_r^2 of 24.64 indicates significant differences among all pairs of rankings, i.e., sections D and I are significantly less valuable than their closest competitor, section H, and so on.

paired and presented to the children who were to choose between the alternatives in each pair. The activities were ranked according to attractiveness. The reliability or stability of the rankings was assessed using a test-retest procedure. The validity of the rankings was determined by a correlational analysis between paired-comparison rankings and actual frequency counts of the same classroom behavior as that used in the paired-comparison presentations, and by a correlation of paired-comparison rankings with rankings on a two-choice task using an apparatus which presented reading or arithmetic materials at the press of a button.

The activity categories were obtained by observing frequently and regularly occurring classroom events. Simple line drawings of each activity were made. The drawings were arranged in all 21 possible pairs, and slides were made of each pair. Slides were shown while a synchronized taped voice asked: "Which of these activities do you do? *Special Activities* (like drawing maps, coloring, or cutting out decorations) or *Checking With The Teacher* (to see whether an answer is right, to find out the assignment or to tell him something interesting)."

The four highest ranking activities for males were *Arithmetic*, *Reading*, *Special Activities*, and *English*, in that order; for females they were *English*, *Special Activities*, *Reading*, and *Arithmetic*, in that order. Stability of highest and lowest paired-comparison choices for a two-week interval was determined for 45 third graders. Agreement was determined as follows: An activity which ranked 1 (highest frequency of the seven activities) on the first administration of the paired comparison task was counted as an agreement in choice two weeks later only if the activity was chosen with sufficient frequency to place it between the ranks of 1 and 3.5. If the activity was ranked 7 on the first administration, agreement two weeks later meant ranking from 3.6 to 7 on the retest. Seventy-six percent of the originally high frequency responses met the criterion; 96 percent of the low-frequency responses met the criterion. Thus, the stability rankings of extreme cases were adequate, particularly for initially low-frequency responses.

Two concurrent validation methods were explored to determine whether the paired-comparison rankings were similar to those obtained by other techniques with apparently greater face validity. The first involved tallying the frequency of the seven activities within a classroom using an Esterline-Angus 20-pen Event Recorder adapted for recording frequency and duration of responses. No significant relationships were found between choices on the paired-comparison

presentation test and actual behavior within the classroom during the free-choice period. The second validation method employed was a correlation of the paired-comparison frequency rankings of activities with the rankings based on a two-choice task. The two-choice task was composed of reading and arithmetic materials. The reading materials were short paragraphs from an SRA *Reading Laboratory* modified to obtain similar duration of response for each selection. Arithmetic materials contained addition and subtraction problems from the ABC *Modern Mathematics Series*, Grade 1. Agreement for reading activity between the two-choice test frequencies and paired-comparison frequencies was 80 percent for males and 92 percent for females. Agreement for the arithmetic activity was much lower. Thus, a simple paired-comparison approach to getting frequency rankings was highly predictive of rankings based on a two-choice task using an apparatus which allowed a choice between arithmetic and reading problems.

The final stage of Dr. Alter's study followed the Premack paradigm. Each child in the experimental group participated in three sessions. The first was a baseline session to determine the child's high-frequency response (arithmetic or reading). The second was a contingency session in which the high-frequency response (arithmetic or reading) could be performed only following performance of the low-frequency response. The third (extinction) session was a return to baseline conditions in which there were no contingency relationships established. A control group also participated in three sessions, which were conducted under baseline or noncontingency conditions.

Subjects were 24 third graders (12 male and 12 female). The design was a 2 (sex) × 2 (experimental-control group) × 2 (high probability reading-high probability arithmetic) × 3 (sessions) factorial with repeated measures on sessions. Each subject had 40 trials within a session. The apparatus used for presenting materials was the two-choice task apparatus referred to above. Under baseline conditions both response buttons were operative and produced stimulus materials (arithmetic or reading) whenever they were pressed. During the contingency session one of the two response buttons was inoperative until the other button was depressed, thus forcing the high-frequency activity to be contingent upon performance of the low-frequency activity.

Major Findings

The major findings of this study were that: Baseline performances were highly stable (control group performance did not differ significantly across sessions I, II, and III); experimental and control groups did not differ significantly under session I baseline conditions, nor did they differ significantly in session III during which both groups were tested again under baseline conditions; low-probability response frequency for experimental subjects was significantly higher in session II than in session I and was higher for the experimental group than for the control group, and the results were the same whether the high-probability response was reading or arithmetic.

Thus, a simple paired-comparison procedure for determining response probabilities was developed. Frequency rankings of activities were found to be highly stable over a two-week period for high- and low-frequency activities. Validity of the paired-comparison rankings was supported by high correlation with frequency of a choice in the two-choice task. Validity of paired-comparison rankings was also supported by an increase in initially low-probability responses produced under conditions in which they were requisites to performing high-probability activities.

Chan's Study

Dr. Chan's study was an outgrowth of Dr. Alter's investigation. While Dr. Alter's work supported Premack's earlier findings, there remained the question of the extent to which the reinforcement effect of high-probability responses was due to *frequency of reinforcement or response rate*. Three studies were conducted to answer this question. Experiment 1 manipulated response rate, while holding reinforcement frequency constant, to evaluate the role of rate alone. Experiment 2 manipulated reinforcement frequency, while holding response rate constant, to evaluate the role of reinforcement frequency alone. Experiment 3 varied both reinforcement frequency and response rate to evaluate the role of both factors simultaneously.

The results of all three experiments suggest that the instructional variable became a contaminating factor. When the experimenter made comments such as "Go faster on this button to get to the side you like," the results clearly yielded rate changes (increases) as a function

of reinforcement frequency and not response rate. But, for minimal cues given to the subjects, no rate change occurred as a function of changes in response rate or reinforcement frequency. Thus, the role of the instructional variable needs to be explored further before unequivocal interpretations can be made of the relative role of *response rate* and *reinforcement frequency* in findings supporting Premack's hypothesis.

APPENDIX A

Scoring Summary

CSE	***Contingent Stimuli Established***
S^{rp}	Scored if the teacher offers and describes a reward for appropriate behavior
S^{rn} *avoid*	Scored if the teacher describes a punishment that will be imposed for inappropriate behavior
S^{rn} *escape*	Scored if the teacher promises to allow his pupils to escape a promised punishment if they behave appropriately
CSA	***Contingent Stimuli Applied***
S^{rp}	Scored if the teacher rewards his pupils as promised
S^{rn} *avoid*	Scored if the teacher punishes his pupils as promised
S^{rn} *escape*	Scored if the teacher allows escape from a punishment he has imposed
RCSA	***Response Contingent Stimuli Applied***
S^{rp}	Scored if the teacher verbally rewards his pupils without first promising reward
ext	Scored if the teacher rewards his pupils with extrinsic reinforcers
tok	Scored if the teacher rewards his pupils with a token (anything eventually traded for a reward)
S^{rn}	Scored if the teacher punishes his pupils without first warning them
T.O.	Scored if the teacher punishes his pupils with time out (isolation)
Comment Spaces:	Provided for observer's comments
Timing:	Used to record observation beginning and ending times
Scoring Responses:	Each time a category is scored, the time the behavior occurred is noted in the proper subcategory spaces

Time Begin *2:10* Time End *2:40*

Sample Observation Record

CSE	S^{rp} 2:14, 2:21, 2:24	COMMENT
	S^{rn} AV 2:23	COMMENT
	S^{rn} ES	COMMENT
CSA	S^{rp} 2:31	COMMENT
	S^{rn} AV	COMMENT
	S^{rn} ES	COMMENT
RCSA	S^{rp} 2:39	COMMENT
	tok ext	
	S^{rn}	COMMENT
	T.O.	

GENERAL COMMENT:

APPENDIX C

General Rating-scale Procedure

Certain teacher responses are listed on the rating scale and are rated according to the code and procedure outlined below. Raters need two sharpened pencils, a clipboard with a stopwatch attached, blank note paper, a rating pack, and a supply of rating sheets.

Rating sheets are divided into nine 30-second intervals on the horizontal axis and five teacher response categories on the vertical axis. Rating packs are made up of pictures of 10 children within a given classroom. The children's names are printed on their pictures.

General Coding Procedure

1. Ratings are to be coded only when the regular teacher is present.
2. Raters are to draw a picture from the top of the shuffled, face-down stack and record or code that child's behavior and the teacher's responses to this child for 4½ minutes, select another, observe for 4½ minutes, and so on until all 10 children have been observed.
3. Raters are to observe during the first 10 seconds of each of the nine 30-second rating periods, within the 4½ minutes for each child, noting which behavior occurs.
4. Raters are to record or code the observed behavior during the final 20 seconds of each 30-second observation interval. They are not to observe during this time.
5. Raters are to record or code all behavior that occurs during an observation interval.
6. If a certain element of behavior occurs more than once during an observation interval, raters are to record or code all of the observations which were noted, unless it is indicated otherwise in the instructions.
7. Raters are not to respond to any child in the classroom but are to ignore the children.
8. Raters are to use only the coding criteria as outlined in the instructions. If an element of behavior cannot be rated according to instructional criteria, note that it cannot be. Do not try to judge behavior or its intent.
9. Raters should be trained in the use of the scale according to the detailed procedures on the training sheet before attempting any data collection.

Student Behavior

Student behavior (SB) is listed as Area 1 and is located on Row 1 of the rating scale.

Rating Procedure

1. Student behavior is to be rated within every 30-second period.
2. The ratings are to be based upon a 4½-minute sample of the target child's behavior. Target children are to be rotated every 4½ minutes.
3. Student behavior is to be listed as either desirable (D) or undesirable (U). If no undesirable behavior occurs during a 30-second rating interval, the interval is coded D. If one or more instances of undesirable behavior occur, that interval is coded U.

Undesirable Behavior Includes:

1. talking aloud without permission;
2. making nonverbal noise such as tapping a pencil on a desk;
3. wandering around the room without instruction from the teacher;
4. disruptive motor behavior such as fighting, wiggling, and poking other children, even if the behavior is instigated by another child;
5. slowly or improperly getting or returning materials;
6. failing to begin, continue, and complete classwork on time as directed;
7. failing to attend during teacher presentations; and
8. leaving the seat without permission, unless regularly permitted to do so.

Nonverbal Teacher Behavior Toward Target Child

Nonverbal teacher behavior toward the target child (NV-T) is listed as Area 2 and is located on Row 2 of the rating scale.

Rating Procedure

1. The rater is to code any nonverbal teacher behavior toward the target child:
   p—if the teacher points at the child;
   c—if the teacher touches or otherwise contacts the child;

h—if the teacher reacts by smiling, winking, nodding, sticking tongue out, frowning, grimacing, head shaking, looking, or any response given with the head; and

a—if the teacher approaches the child, touches his desk or materials thereon, but does not touch him.

2. In addition, a plus sign (+) is added to any of the above codings when the teacher's behavior is unquestionably approving, and a minus sign (—) when the teacher's behavior is unquestionably disapproving. Disregard the plus (+) and minus (—) signs if in doubt.

Verbal Teacher Behavior Toward Target Child

Verbal teacher behavior toward the target child (V-T) is listed as Area 3 and located on Row 3 of the rating scale.

Rating Procedure

1. The rater is to code any verbal teacher behavior toward the target child:

(+)—if the teacher states that the target child is engaging in a D behavior, is not engaging in a U behavior, is to receive some positive reinforcement, or otherwise praises him;

(—)—if the teacher states that the child is engaging in a U behavior, is not engaging in a D behavior, is to receive something aversive, or otherwise reprimands or criticizes him;

I—if the teacher gives an assignment, answers a child's question, indicates what the child is to do or how he is to do it, or otherwise instructs him;

O—if the teacher verbally interacts with the child in a way not clearly part of another code.

2. If the teacher specifies another child or children *along with* the target child, rate the teacher's verbal behavior in row V-T. The target child may be the only child spoken to or he may be specified by name along with other children. A rating is not made in row V-T if the teacher does not in some way specify the target child as his spoken target while excluding most of the others in the class.
3. Note that in row V-T more than one code can often be recorded. For instance, if the teacher instructs the target child and praises

him in addition, the rating becomes I+. An example of this would be, "Johnny, when you finish reading, you may go to recess." A + only code does not include an instruction; e.g., "Johnny, you've worked so hard today that you may go to recess early." Coding combinations are similarly used with the minus sign.

Verbal Teacher Behavior Toward Other Than Target Child

Verbal teacher behavior toward others (V-O) is listed as Area 4 and is located on Row 4 of the rating scale.

Rating Procedure

The rater is to code verbal teacher behavior which is in no way directed towards the target child. The rater should note whether the teacher specifies another child or children or whether she directs her statement to the class in general.

The codes and procedures used for Area V-T are also applicable in Area V-O.

General Character of Teacher Interaction

General character of teacher interaction (TI) is listed as Area 5 and is located on Row 5 of the rating scale. Interactions may include questions, statements, explanations, prompts, probes, calling on a child, etc.; and, depending upon the teacher's intent, any of these can be academic instruction, schedule instruction, or behavior management.

Rating Procedure

1. The rater is to code at least one type of teacher interaction within every rating interval:
   i—if the teacher interacts with one student;
   g—if the teacher interacts with a group of students ranging from two children to approximately one-half of the class; and,
   c—if the teacher interacts with more than one-half of the class.
2. The teacher may work with a single individual as well as speak to the class during a single 10-second interval; therefore, there is a

good possibility that all three codes may be used during any one rating interval.

3. Additional code specifications are used in conjunction with the above when:
   (a) The interaction is basically academic. If the interaction concerns academic work or content, circle the i, g, or c. Examples are: "6 + 8 are 14," "Your answer is correct," or "I am sure you remember who saw the bunny."
   (b) The interaction basically concerns scheduling. If the interaction concerns changing or moving activities, locations, materials, etc., as a function of the class schedule, prime the i, g, or c. Examples are: "Put your papers away now," "The time is nearly up," or "Let's sit at the large table."
   (c) The interaction basically concerns behavior management. If an interaction is an attempt by the teacher to get a child to stop emitting a U behavior or an attempt to get a child to emit a D behavior, underline it. Examples are "Turn around in your seats," "Be quiet."
   (d) If an interaction cannot be coded as academic, scheduling, or behavior management, code it i, g, or c.

Summary

AREA 1 *Student Behavior*

U (undesirable) or D (desirable), (score one—give U preference)

AREA 2 *Nonverbal Teacher Behavior Toward Target Child*

p (points), c (contact), h (head), a (approaches), (score all)
Score + or — if appropriate

AREA 3 *Verbal Teacher Behavior Toward Target Child*
(score applicable behavior)

+ positive
— negative
I instruction
O other verbalization

AREA 4 *Verbal Teacher Behavior Toward Other than Target Child*
(score all)

+ positive
− negative
I instruction
O other verbalization

AREA 5 *General Character of Teacher Interaction*
(score applicable behavior)

i individual academic interaction
g group academic interaction (less than half the class)
c class academic interaction

i′ individual schedule instruction
g′ group schedule instruction (less than half the class)
c′ class schedule instruction

i individual behavior management
g group behavior management (less than half the class)
c class behavior management

i individual interaction, other
g group interaction, other
c class interaction, other

APPENDIX D

PLEASE READ THE DIRECTIONS CAREFULLY BEFORE ATTEMPTING TO COMPLETE AN ITEM.

I.
Please check (✓) the appropriate boxes:

	Yes	No
Do you read the ECRI Newsletter?	☐	☐
Would you like to continue receiving future issues?	☐	☐

If you answer **No**, we must have your name and address so that we can delete it from our Newsletter mailing list. See item VI. below.

III.

Please rate **each** Newsletter Section (rows A through I) on **each** Rating Category (columns 4 through 12). Using the five-point scale, fill in **each** box with the number which best represents your view of how clear, informative, important, etc., **each** Newsletter Section is:

1	2	3	4	5
excellent	above average	average	below average	poor

II.

Please check (✓) the appropriate boxes in Columns 1 through 3:

NEWSLETTER SECTIONS	Sections which you read regularly	Sections which you would give more space in the Newsletter	Sections which are of no value to you	Influential	New, original, innovative	Practical, sound, realistic	Applicable	Useful, valuable	Important, significant	Interesting, challenging	Informative, instructive	Clear, easily understood
	1	2	3	RATING CATEGORIES								RATING CATEGORIES
A. Feature article (front page)												
B. About the author												
C. Exemplary teaching practices												
D. Letters to the editor												
E. ECRI sponsored projects												
F. Reading research review												
G. Reading keys												
H. Cartoons												
I. Photographs												

IV.

Please check (✓) the box which most accurately represents your current job category. If you have to use a box marked **Other**, please specify your current position in the appropriate box:

	Teacher	Administrator	Other
Elementary			
Secondary			
Jr. College			
College			
Other			

V.
Comments: _____

VI.
Name and address of evaluator (optional unless answer to question two, item I is No):

Name _____ Address _____ Zip _____